
Aprendizagem Automática

FICHA N. 4

ENUNCIADO

Nome: Fábio Alexandre Cruz Silva Dias

Número: A42921

ATENÇÃO: Fixa de respostas múltiplas. Só uma única resposta em cada alínea está correta. Cada alínea vale 2 valores. Respostas erradas descontam 0.5 valores.

1. No ficheiro `A42921_Q001_data.p` encontra-se disponível uma variável independente x e uma variável dependente y . Pretende-se estimar a variável y através de uma regressão polinomial da variável x , minimizando o erro quadrático médio.
 - (a) Considere o conjunto de treino composto pelo “fold” 1, e o conjunto de teste composto pelo “fold” 0. Considere ainda que, através da minimização do erro quadrático médio do conjunto de treino, se estimou um modelo regressão polinomial de 4ª ordem: $\hat{y} = w_0 + w_1x + w_2x^2 + w_3x^3 + w_4x^4$.
 - i. No conjunto de treino, o coeficiente de determinação, R^2 , é igual a 0.40.
 - ii. No conjunto de teste, o coeficiente de determinação, R^2 , é igual a 0.30.
 - iii. Todas as respostas anteriores.
 - iv. Nenhuma das respostas anteriores.
 - (b) Considere o conjunto de treino composto pelo “fold” 0, e o conjunto de teste composto pelo “fold” 1. Considere ainda que, através da minimização do erro quadrático médio do conjunto de treino, se estimou um modelo regressão polinomial de 3ª ordem: $\hat{y} = w_0 + w_1x + w_2x^2 + w_3x^3$.
 - i. No conjunto de treino, o erro absoluto médio é igual a 2.27.
 - ii. Arredondando a zero casas decimais, o valor de w_0 é -44.
 - iii. Todas as respostas anteriores.
 - iv. Nenhuma das respostas anteriores.
2. Considere o conjunto “diabetes” disponível em `sklearn.datasets` (usar a função `load_diabetes()`). Pretende-se estimar e avaliar modelos de regressão polinomial com os dados deste conjunto: use as primeiras 295 amostras para treino e as restantes para teste.
 - (a) Considere que utiliza função Lasso (sub-módulo `linear_model` do `sklearn`) para uma regressão polinomial de 4ª ordem dos dados de treino. Instancie o regressor somente com os seguintes parâmetros:
`Lasso(random_state=42, alpha=0.01)`
 - i. No conjunto de treino, o coeficiente de determinação, R^2 , é igual a 0.54.
 - ii. Excluindo w_0 , o número de coeficientes do polinómio com valor absoluto maior que 212 é igual a 11.
 - iii. Todas as respostas anteriores.

- iv. Nenhuma das respostas anteriores.
 - (b)
 - i. O número de coeficientes, incluindo w_0 , numa regressão polinomial de ordem 3, é igual a 286.
 - ii. O número de coeficientes, incluindo w_0 , numa regressão polinomial de ordem 4, é igual a 1001.
 - iii. Todas as respostas anteriores.
 - iv. Nenhuma das respostas anteriores.
 - (c) Considere o modelo de regressão polinomial de 2ª ordem que minimiza o erro quadrático médio no conjunto de treino.
 - i. No conjunto de teste, o erro quadrático médio é igual a 8891.35.
 - ii. No conjunto de treino, o erro quadrático médio é igual a 2271.42.
 - iii. Todas as respostas anteriores.
 - iv. Nenhuma das respostas anteriores.
3. No ficheiro `A42921_Q003_data.p`, encontra-se um conjunto de dados bidimensionais divididos em duas classes $\Omega = \{\varpi_0, \varpi_1\}$ (negativos e positivos). Há duas variáveis num dicionário: \mathbf{x} é uma matriz de dados, e \mathbf{y} é um array com as classes dos dados. Considere o seguinte modelo linear de classificação:
- $$\hat{y} = \mathbf{w}^\top \mathbf{x} = w_0 + w_1 x_1 + w_2 x_2, \text{ com } \mathbf{x} \in \varpi_1 \text{ para } \hat{y} \geq 0, \text{ e para } \mathbf{w} = [w_0, w_1, w_2]^\top \text{ e } \mathbf{x} = [1, x_1, x_2]^\top.$$
- Considere ainda que o vetor \mathbf{w}_{MSE} é o vetor de pesos que minimiza o erro quadrático médio deste conjunto: $\mathcal{E} = \frac{1}{N} \sum_{n=1}^N (y[n] - \hat{y}[n])^2$, onde N é o número total de pontos, e $n = 1, \dots, N$. As saídas desejadas são: $y[n] = -1$ para $\mathbf{x}[n] \in \varpi_0$ e $y[n] = +1$ para $\mathbf{x}[n] \in \varpi_1$.
- (a) Consider o classificador com o seguinte vetor de pesos $\mathbf{w} = [0.00, 0.95, -0.31]$.
 - i. O número total de acertos é de 1936.
 - ii. O valor do *recall* é de 0.896.
 - iii. O número de erros na classe ϖ_0 é de 84.
 - iv. O número de erros na classe ϖ_1 é de 145.
 - (b) Considere o classificador com vetor de pesos, \mathbf{w}_{MSE} , que minimiza o erro quadrático médio do conjunto.
 - i. O erro absoluto médio é igual a 0.612.
 - ii. O erro quadrático médio é igual a 0.476.
 - iii. Todas as respostas anteriores.
 - iv. Nenhuma das respostas anteriores.
 - (c) Consider o classificador com o seguinte vetor de pesos $\mathbf{w} = [0.00, 0.49, 0.87]$.
 - i. O número de erros na classe ϖ_0 é de 239.
 - ii. O valor do *recall* é de 0.826.
 - iii. Todas as respostas anteriores.
 - iv. Nenhuma das respostas anteriores.